Int. J. Metrol. Qual. Eng. **3**, 117–123 (2012)
© EDP Sciences 2012
DOI: 10.1051/ijmqe/2012010

Least-squares fitting with errors in the response and predictor

T. Burr[1,*], S. Croft[2], and B.C. Reed[3]

[1] Statistical Sciences, Los Alamos National Laboratory, USA
[2] Safeguards Science and Technology, Los Alamos National Laboratory, USA
[3] Department of Physics, Alma College, USA

Received: 10 August 2011 / Accepted: 22 March 2012

Abstract. Least squares regression is commonly used in metrology for calibration and estimation. In regression relating a response y to a predictor x, the predictor x is often measured with error that is ignored in analysis. Practitioners wondering how to proceed when x has non-negligible error face a daunting literature, with a wide range of notation, assumptions, and approaches. For the model $y_{\text{true}} = \beta_0 + \beta_1 x_{\text{true}}$, we provide simple expressions for errors in predictors (EIP) estimators $\hat{\beta}_{0,\text{EIP}}$ for β_0 and $\hat{\beta}_{1,\text{EIP}}$ for β_1 and for an approximation to covariance $(\hat{\beta}_{0,\text{EIP}}, \hat{\beta}_{1,\text{EIP}})$. It is assumed that there are measured data $x = x_{\text{true}} + e_x$, and $y = y_{\text{true}} + e_y$ with errors e_x in x and e_y in y and the variances of the errors e_x and e_y are allowed to depend on x_{true} and y_{true}, respectively. This paper also investigates the accuracy of the estimated $\text{cov}(\hat{\beta}_{0,\text{EIP}}, \hat{\beta}_{1,\text{EIP}})$ and provides a numerical Bayesian alternative using Markov Chain Monte Carlo, which is recommended particularly for small sample sizes where the approximate expression is shown to have lower accuracy than desired.

Keywords: Least square; regression; Bayesian estimation; errors

1 Introduction

Least squares regression is commonly used in metrology for calibration and estimation. In this paper, the assumed model relating the true response y_{true} to the true predictor x_{true} is

$$y_{\text{true}} = \beta_0 + \beta_1 x_{\text{true}}. \quad (1)$$

The measured response y is modeled as

$$y = y_{\text{true}} + e_y, \quad (2)$$

and $x = x_{\text{true}} + e_x$ is the measured predictor.

Using n pairs of (x, y) values, we seek errors-in-predictors (EIP) estimates $\hat{\beta}_{0,\text{EIP}}$ for β_0 and $\hat{\beta}_{1,\text{EIP}}$ for β_1 in situations for which β_0 and β_1 are of interest, not merely for their role in using x to predict y_{true}. In addition, some applications seek an estimate of y_{true} using both x and y and in this context β_0 and β_1 are of not of interest for their own sake, but the EIP estimates $\hat{\beta}_{0,\text{EIP}}$ and $\hat{\beta}_{1,\text{EIP}}$ can be used to develop an optimal estimate of y_{true}, where "optimal" is defined as the estimator having smallest average squared error around y_{true}.

Denote the standard deviations of y and of x as σ_{y_i} and σ_{x_i}, respectively, where the notation is chosen to allow for the y_{meas} and/or x_{meas} standard deviations to depend on the respective true values. For compact notation below, X denotes x_{true} and Y denotes y_{true}.

* Correspondence: tburr@lanl.gov

It is fairly widely known that ordinary least squares (OLS) (or weighted least squares (WLS) if σ_{y_i} depends on y_i) estimates of β_0 and β_1, which ignore the errors in x are biased due to the observation errors in x_{meas} [1–7]. However, it appears to be less widely known that if the main inference goal is prediction, using future x to predict future Y, then bias in OLS estimates is of little or no concern. This is because the bias in OLS estimates $\hat{\beta}_0$ for β_0 and $\hat{\beta}_1$ for β_1 does no harm when x rather than x_{true} is used for prediction, assuming future x have the same error distribution as the training x used to estimate β_0 and β_1 [1]. So, if the main inference goal is to use future x to predict future Y, then usual weighted least squares estimators $\hat{\beta}_{0,\text{WLS}}$ and $\hat{\beta}_{1,\text{WLS}}$ ignoring errors in x are appropriate. However, if the goal is to use both x and y to estimate y_{true} or to estimate β_0 and/or for β_1, then error in x should not be ignored. See Appendix A.3.

If one seeks unbiased estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, then an EIP alternative to WLS is needed and a few have been provided [3–6, 8, 9] using various approximations. Regarding notation, unless stated otherwise, $\hat{\beta}_0$ and $\hat{\beta}_1$ will denote the EIP-based estimators. This paper provides simple expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$, and for $\text{cov}(\hat{\beta}_0, \hat{\beta}_1)$, allows σ_{y_i} and σ_{x_i} to depend on the respective true and provides a numerical Bayesian alternative using Markov Chain Monte Carlo (MCMC) which is recommended particularly for small sample sizes where the approximate expression is shown to have lower accuracy than desired.

Article published by EDP Sciences

https://doi.org/10.1051/ijmqe/2012010 Published online by Cambridge University Press

Several studies deal with challenges and limitations of EIP approaches. For example, Carroll et al. [1] show that if model error is present, then EIP estimators overcorrect for OLS bias. Brown [8] develops M estimators which are intended to be more robust to outliers than ordinary ML or moments-based estimators. Fuller [3] considers cases in which σ_{x_i} and/or σ_{y_i} are not known exactly but must be estimated based on specified sample sizes. However, Fuller's EIP estimators [3] aim to have minimum mean squared error (MSE), derived from asymptotic arguments. Performance in small data sets is only briefly studied, with indications that the corresponding estimated covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$ is noticeably different from the true covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$.

This paper is organized as follows. Section 2 provides EIP estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, then uses a result that follows from simple algebra to provide an estimate of the covariance matrix $\text{cov}(\hat{\beta}_0, \hat{\beta}_1)$ used in Section 3. Section 4 gives an example in which equation (1) describes two measurement methods of the same underlying true quantity, so x and y are both "responses" to the true value. In this example, EIP is recommended as a way to improve the estimated true measurand value by using both x and y. Section 5 introduces a numerical Bayesian option for EIP. Section 6 is a summary.

2 EIP estimators $\hat{\beta}_0$ of β_0 and $\hat{\beta}_1$ of β_1

Assuming Gaussian errors in x and y, it is easily shown [8] that the maximum likelihood (ML) based EIP estimators of β_0, β_1, and X_i are those values minimizing the sum of the weighted squared residuals, designated as S:

$$S = \sum_{i=1}^{n} \left[W(x_i)\ (x_i - X_i)^2 + W(y_i)\ (y_i - Y_i)^2 \right], \quad (3)$$

where (x_i, y_i) are the measured values of the n data pairs and $Y = \beta_0 + \beta_1 X$. The $W(x_i)$ and $W(y_i)$ are the x- and y- weights of each point, which are typically taken to be, based on ML arguments, the inverse-squares of their standard deviations σ_{x_i} and σ_{y_i}, $W(x_i) = 1/\sigma_{x_i}^2$ and $W(y_i) = 1/\sigma_{y_i}^2$. If the errors in x and y are non-Gaussian, defensible estimates of β_0, β_1, and X_i can still be based on minimizing S.

We note that some treatments [10] use what might appear to be a different optimization criterion, choosing to minimize

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma_{y_i}^2 + \beta_1^2 \sigma_{x_i}^2} \right], \quad (4)$$

where $\sigma_y^2 + \beta_1^2 \sigma_x^2$ is the variance of $(y_i - \beta_0 - \beta_1 x_i)$, so χ^2 has a chi-squared distribution if the errors in x and y have a Gaussian distribution, and approximately so if the errors in x and y are not badly non-Gaussian or if n is large. It is simple to show (see below) that estimation on the basis of minimizing S gives the same estimators as estimation on the basis of minimizing χ^2.

An estimation approach that traces to work first developed by [2] allows for a simple expression for an estimator of $\text{cov}(\hat{\beta}_0, \hat{\beta}_1)$ as follows. ML estimation leads to estimating X values [9] extended here to allow for nonzero β_0 as

$$\hat{X}_i = x_i + \frac{\beta_1 (y - \beta_0 - \beta_1 x_i)}{\beta_1^2 + \lambda_i}, \quad (5)$$

where $\lambda_i = \frac{\sigma_{y_i}^2}{\sigma_{x_i}^2}$. Define an "overall weight" for each data pair as

$$W_i = \frac{W(x_i)\ W(y_i)}{\beta_1^2 W(y_i) + W(x_i)} = \frac{1}{\beta_1^2 \sigma_{x_i}^2 + \sigma_{y_i}^2}. \quad (6)$$

Notice that the weights W_i correspond to the weights in χ^2 in (4). Using equation (6) and substituting equation (5) into equation (3), by simple algebra, the sum of squared residuals in equation (1) can be cast into a form reminiscent of that used for the traditional errors-in-y-only formulation and identical to (4):

$$S = \sum_{i=1}^{N} W_i\ (y_i - \beta_0 - \beta_1 x_i)^2. \quad (7)$$

The weights in equation (4) were defined by calculating the variance of $(y_i - \beta_0 - \beta_1 x_i)$. The same weights in equation (7) arose from a more complicated ML approach that also provides equation (5) which is needed in Section 4.

From equation (4) or (Eq. (7)), and using matrix notation, the usual WLS estimate $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ is given by

$$\hat{\beta} = (X^T W X)^{-1} X^T W y \quad (8)$$

where the diagonal matrix W has values W_i along its diagonal and X is a 2-column matrix with 1s in column 1 and x_is in column 2.

It then follows (as a standard result for weighted least squares regression [10]), that the covariance matrix Σ for $\hat{\beta}_0$ and $\hat{\beta}_1$ can be written as

$$\Sigma = (X^T W X)^{-1}. \quad (9)$$

Remark 1. Equations (8) and (9) include the unknown β_1. Therefore, any EIP approach via equations (8) and (9) is not identical to WLS because this EIP approach requires an iterative procedure. However, we emphasize the connection to WLS and equation (9) in particular which leads to a reasonable option to estimate the covariance matrix Σ for $\hat{\beta}_0$ and $\hat{\beta}_1$. Regarding equation (8), although we do not think an iterative procedure for this EIP approach is available to estimate both β_0 and β_1, we did for completeness here verify that a simple 1-parameter-search iterative procedure leads to the value of β_1 that minimizes S in equation (7) (or equivalently that minimizes χ^2 in Eq. (3)). The iterative procedure requires mean-centering as in Appendix A.1 so that $\beta_0 = 0$ can be assumed during the iterative search for β_1. Once the value of β_1 is found that minimizes S in equation (7),

https://doi.org/10.1051/ijmqe/2012010 Published online by Cambridge University Press

$$\hat{\sigma}_{\beta_1} \approx \sqrt{\sum_{i=1}^{n}\left[\frac{1}{W(x_i)}\left(\frac{\partial\beta_1}{\partial x_i}\right)^2 + \frac{1}{W(y_i)}\left(\frac{\partial\beta_1}{\partial y_i}\right)^2 + \left(\frac{\partial\beta_1}{\partial x_i}\right)\left(\frac{\partial\beta_1}{\partial y_i}\right)\mathrm{cov}(x_i, y_i)\right]}, \tag{10}$$

it is easily shown that $\hat{\beta}_0 = \langle y\rangle - \hat{\beta}_1\langle x\rangle$, which has the same form as arises in the WLS case without error in x, and $\langle x\rangle$ and $\langle y\rangle$ are defined in Appendix A.1.

Alternatively, to find $\hat{\beta}_0$ and $\hat{\beta}_1$ that together minimize S in equation (3), for the case $\sigma_{x_i} = \sigma_x$ and $\sigma_{y_i} = \sigma_y$, Madansky [9] showed how to unwrap this implicit characterization without using iteration, resulting in the expression

$$\hat{\beta}_1 = \frac{\delta + \sqrt{\delta^2 + 4\lambda s_{xy}}}{2s_{xy}},$$

where $\delta = s_{yy} - \lambda s_{xx}^2$,

$$s_{xx}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2, s_{yy}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2,$$

and

$$s_{xy}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

And, Reed [5] provides the analogous result (Appendix A.1) for the more general case with σ_{x_i} and σ_{y_i} depending on the true value for observation i. We have verified that Reed's [5] more general expression gives the same estimators as ML as implemented for example using optim in R [12] which can optimize over multiple parameters simultaneously. Here, either a 2-parameter or a 1-parameter optimizer will suffice. If a 1-parameter optimizer is used to find $\hat{\beta}_1$, the relation $\hat{\beta}_0 = \langle y\rangle - \hat{\beta}_1\langle x\rangle$ must be enforced at each stage. Initial estimates should of course be reasonable, such as those by using WLS that ignores error in xs by assuming $\sigma_{x_i}^2 = 0$ for each x_i.

Remark 2. There are several existing EIP approaches depending on what is known about $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$. For example, if only the ratio $\lambda_i = \frac{\sigma_{y_i}^2}{\sigma_{x_i}^2}$ is known, then equation (9) might be modified by including the factor $S/(n\text{-}2)$, where S is given in equation (7). For a good discussion of assigning weights we suggest [11], which points out that if the fitting model (Eq. (1) in the present case) is "exact" and hence that the subsequent deviations result purely from imprecisely measured experimental values, then the inverse-square approach is entirely appropriate.

Remark 3. It is simple to show from equation (7) that the ML estimate of β_0 is

$$\hat{\beta}_0 = \langle y\rangle - \hat{\beta}_1\langle x\rangle,$$

where

$$\langle y\rangle = \frac{\sum_{i=1}^{n} W_i y_i}{\sum_{i=1}^{n} W_i} \text{ and } \langle x\rangle = \frac{\sum_{i=1}^{n} W_i x_i}{\sum_{i=1}^{n} W_i}$$

which is the natural extension of the familiar OLS result $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ which uses the ordinary unweighted average x and y values.

Remark 4. The main reason for making the connection to WLS is that equation (9) can be shown to provide the same estimate of the covariance matrix Σ for $\hat{\beta}_0$ and $\hat{\beta}_1$ as has been provided in [5] using tedious algebra and delta method approximations (Appendix A.3).

So, the key approximate result here is equation (9), which is simpler than analogous approximations in Fuller [3] and in Reed [5], and which is to be evaluated at observed xs.

Fuller [3] used approximations to develop estimators having asymptotically minimum MSE, where MSE = variance + bias^2. Note however that if each X_i, β_0 and β_1 were known exactly, then equation (9) would be exact for $\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1)$. However, because approximations are involved, it is not clear for estimation performance reasons whether equation (9) should be evaluated at the observed x values or the estimated X values given in equation (5). And, unlike WLS or OLS which lead to exact expressions for $\mathrm{cov}(\hat{\beta}_{0,\mathrm{WLS}}, \hat{\beta}_{1,\mathrm{WLS}})$, the EIP based expressions for all analytical approaches such as presented in this section involve approximations. Note also that the weights W_i in equation (6) assume that both $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$ are known. It is perhaps more typical that the x and y variances be expressed as a function of the underlying true x_{true} and y_{true} values, respectively, which are unknown. Then because the estimates $\hat{X}_i$ and $\hat{y}_i$ provide better estimates of x_{true} and y_{true}, it could be argued that the x and y variances should be based on $\hat{X}_i$ and $\hat{y}_i$. Because this paper assumes zero uncertainty in $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$, this issue is not considered here and in the results given below, equation (9) is evaluated at the observed (measured) x values. We note that the numerical Bayesian approach via MCMC is probably best equipped to accommodate uncertainty in $\sigma_{x_i}^2$ and/or $\sigma_{y_i}^2$, and can provide essentially an exact result for $\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1)$, but a prior distribution for parameter values is required.

3 Estimation of $\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1)$ for EIP estimators

From Section 2 it follows that there are at least three reasonable options to estimate $\mathrm{cov}(\hat{\beta}_0,)$ for EIP estimators.

Option 1: use equation (9) evaluated at the observed (measured) xs, with the ML estimate $\hat{\beta}_1$ used in the weight matrix W.

Option 2: use well-known asymptotic results involving Fisher information associated with ML (see Appendix). The Fisher information involves derivatives of the likelihood expression with respect to the parameters β_0 and β_1.

Option 3: use the "delta method", described for $\hat{\sigma}_{\beta_1}$ as

see equation (10) above.

https://doi.org/10.1051/ijmqe/2012010 Published online by Cambridge University Press

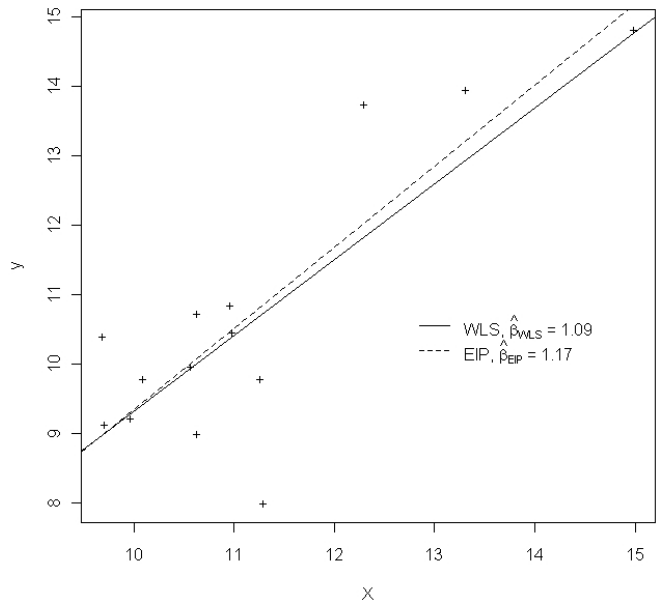


Fig. 1. The example data y versus x for the 14 (x, y) pairs, and lines depicting the WLS and EIP intercept and slope estimates.

which can be derived by expressing the solution for $\hat{\beta}_1$ as a first term (linear) Taylor series around β_1. Note that option 2 involve derivatives with respect to β_0 and β_1 in the likelihood expression, while option 3 involves derivatives with respect to the measured values in an expression for the estimators (we use Reeds [4] expression. See Appendix A.1).

Example 1 is 14 (x, y) pairs [4], with $x_1, x_2, \ldots, x_{14} =$ 13.32, 14.99, 12.30, 10.64, 10.10, 11.27, 9.69, 11.30, 10.99, 10.64, 10.57, 10.97, 9.97, 9.71 and $y_1, y_2, \ldots, y_{14} =$ 13.96, 14.83, 13.75, 9.00, 9.80, 9.80, 10.40, 8.00, 10.47, 10.73, 9.97, 10.86, 9.23, 9.14. We first follow [4] and estimate $(\hat{\beta}_0, \hat{\beta}_1) = (-2.313, 1.166)$. The x and y variances are assumed to be known, with $\sigma_x =$ 0.17, 0.24, 0.50, 0.70, 0.29, 0.70, 0.20, 0.90, 0.60, 0.50, 0.50, 0.50, 0.20, 0.18 and $\sigma_y =$ 1.15, 1.17, 2.80, 1.00, 0.70, 1.00, 1.40, 1.50, 1.50, 1.90, 1.00, 1.60, 0.60, 0.60.

As discussed in Burr and Knepper [13], the bias in the OLS (or WLS) estimate of β_1 depends on the ratio $\frac{\sigma_x^2}{\sigma_x^2+\mathrm{var}(X)}$ which is the attenuation coefficient. The average of $\frac{\sigma_x^2}{\sigma_x^2+\mathrm{var}(X)}$ across the 14 x values is 0.11, and the value 0.10 is often given as a thumb rule for when EIP can be ignored and OLS or WLS used. That is, if the measurement error variance in x is small compared to the variance of the measured xs (which includes the variance of the true Xs and measurement error variance in x), then EIP-based estimates will not differ much from OLS or WLS estimate.

Figure 1 plots y versus x and shows the WLS and EIP estimates of β_0 and β_1. Indeed, only a modest attenuation effect is evident in the WLS estimate. However, in general, EIP estimates and the associated inferences regarding $(\hat{\beta}_0, \hat{\beta}_1)$ can be quite different from WLS

Table 1. Observed and average of the estimated $\{\mathrm{var}(\hat{\beta}_0)$, $\mathrm{var}(\hat{\beta}_1)$, and $\mathrm{cov}(\hat{\beta}_0,\hat{\beta}_1)\}$ for each of three options. Entries are repeatable to ±1% relative or smaller.

	Observed			Average estimated		
Option 1	5.3	0.05	–0.48	4.7	0.04	–0.49
Option 2	5.3	0.05	–0.48	5.0	0.04	–0.46
Option 3	5.3	0.05	–0.49	4.7	0.04	–0.43

estimates. To assess the quality of options 1−3, many data realizations were simulated as follows. Assume the $(\hat{\beta}_0, \hat{\beta}_1)$ values are the true β_0 and β_1 and also assume that the x values are the true values X. That is, generate $Y = 2.313 - 1.166X$ and then apply equation (2) to generate $y = \beta_0 + \beta_1 x_{\mathrm{true}} + e_y$ and let $x = x_{\mathrm{true}} + e_x$. Then generate many realizations of 14 synthetic (x, y) pairs to assess the quality of options 1−3. Using 100 000 simulations in R, the observed and average (over the 100 000 simulations) estimated covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$ are given in Table 1 for options 1−3 (the 3 listed values are $\mathrm{var}(\hat{\beta}_0)$, $\mathrm{var}(\hat{\beta}_1)$, and $\mathrm{cov}(\hat{\beta}_0,\hat{\beta}_1)$). In this example, all three options perform quite similarly, and appear to provide adequate approximations.

In addition to the quality of the estimate of the covariance matrix of $(\hat{\beta}_0, \hat{\beta}_1)$, the RMSE of the estimators $(\hat{\beta}_0, \hat{\beta}_1)$ is of interest. As noted in [1], there have been surprisingly few published studies of RMSE performance. The estimated RMSEs of $(\hat{\beta}_0, \hat{\beta}_1)$ are: (2.4,0.22); (2.3, 0.21), and (2.3, 0.22), respectively for options 1,2, and 3. Again, all options perform acceptably well in this example.

4 Prediction

Recall that options to accommodate errors in predictors should be considered when there is interest in the true β_0 and β_1. In Example 1, there could be interest in whether $\beta_0 = 0$ and $\beta_1 = 1$, which would provide insight regarding relation between the two temperature measurement options [7]. In an example in [14], β_1 is the parameter of interest. Examples in nondestructive assay are given in [13] where the true values of the βs are of interest. Also, there could be interest in a meta-analysis using both y and x to give a better estimate of the true y value, Y, which in Example 1 is the true temperature. Note that if future xs (not yet observed) will be used to predict future Ys, then WLS has lower RMSE than EIP [1], which we verified using data simulated according to the description above for Example 1.

Now consider using the $n = 14$ measured (X, Y) pairs to predict Y in a type of meta-analysis. In this case, EIP performs better than WLS, with an average RMSE of 0.63 for EIP compared to 0.66 for WLS (both repeatable across sets of 10 000 simulations to within ±1% relative). We believe this type of meta-analysis deserves further study.

A second aspect of prediction is the quality of the associated confidence interval (CI) for the predictions. CI quality is defined by how close the nominal coverage is to the actual coverage. For example, a nominal 95% CI

https://doi.org/10.1051/ijmqe/2012010 Published online by Cambridge University Press

should include ("cover") the true parameter in approximately 95% of hypothetical repeats of the data collection and CI construction. Analogously, for the Bayesian strategy described below, the 95% credible interval constructed from the parameter's posterior distribution should have close agreement between actual and nominal coverage.

Each of the 14 predicted Y values is given by $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\hat{X}$ with $\hat{X}$ given by equation (5) using the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. However, we find in Example 1 that the estimated standard deviation of the predicted Ys is consistently different from the observed standard deviation, suggesting either that the Table 1 entries of estimated $\{\text{var}(\hat{\beta}_0), \text{var}(\hat{\beta}_1), \text{and } \text{cov}(\hat{\beta}_0, \hat{\beta}_1)\}$ are not sufficiently accurate, and/or that the estimated variance of $\hat{X}$ is not sufficiently accurate. For example, the predicted average RMSE of $\hat{Y}$ on the basis of using

$$\sigma_{\hat{Y}} \approx \{\sigma^2_{\hat{\beta}_0} + \beta_1^2\sigma^2_{\hat{X}} + \hat{X}^2\sigma^2_{\hat{\beta}_1} + 2\hat{X}\,\text{cov}(\hat{\beta}_0, \hat{\beta}_1) + 2\hat{\beta}_1\text{cov}(\hat{X}, \hat{\beta}_0) + 2\hat{\beta}_0\text{cov}(\hat{X}, \hat{\beta}_1)\}^{1/2}$$

across the 14 true Y values is 0.55, which is approximately 13% less than the observed 0.63 for the EIP-based average RMSE. Note that 13% is outside the repeatability across sets of 10 000 simulations of ±1% relative, so there is non-negligible approximation error in the above expression for σ_y. To evaluate the approximation for $\sigma_{\hat{Y}}$, equation (5) can be used to approximate the terms involving $\hat{X}$. For example, from equation (5) it follows that $\sigma^2_{\hat{X}} = \frac{1}{\beta_1^2/\sigma_x^2 + 1/\sigma_y^2}$ as given in [3].

We note here that generally as sample size increases, the estimate of σ_y should improve. To check this, we augmented the 14 (x, y) pairs with 14 smaller-valued xs (each equal to the original xs – 10) and 14 larger-values xs (each equal to the original xs + 10), for a total of 42 (x, y) pairs. The true values Y were again set to $Y = 2.313 - 1.166X$. The result was a predicted average RMSE of 0.62, compared to a 0.61 observed RMSE, which indicates that a sample size of 42 is sufficiently large for good agreement between predicted and observed.

5 Numerical Bayesian alternative

Because of the relatively large disagreement between the estimated and observed average RMSE in $\hat{Y}$ in the case of $n = 14$ (x, y) pairs, we implemented a numerical Bayesian strategy that relies on MCMC [15] using the likelihood as in equation (3) to get estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{X}$. This strategy does not yet appear in any literature to our knowledge, although [1] mentions the possibility of numerical Bayesian options and Dellaportas and Stephens [16] demonstrate MCMC using their implementation of the Gibbs sampler [15].

MCMC-based Bayesian inference has at least three advantages over the ML approach. First, MCMC provides an almost exact numerical result (to within numerical accuracy which is a function of the number of observations in the Markov Chain) for the estimated $\{\text{var}(\hat{\beta}_0), \text{var}(\hat{\beta}_1), \text{and } \text{cov}(\hat{\beta}_0, \hat{\beta}_1)\}$ and for any other function such as $\hat{Y}$. Second, any likelihood function is easily accommodated. In contrast, expressions for the estimated $\{\text{var}(\hat{\beta}_0), \text{var}(\hat{\beta}_1), \text{and } \text{cov}(\hat{\beta}_0, \hat{\beta}_1)\}$ of ML estimates can be tedious to derive for some likelihoods (see Appendix A.2), and ML via numerical optimization can have convergence difficulties. Third, in most real applications, there is measurement error in σ_x and/or σ_y, which can be accommodated with minor extension to the likelihood for the MCMC approach, but is not so easily accommodated for analytical approaches such as in [3].

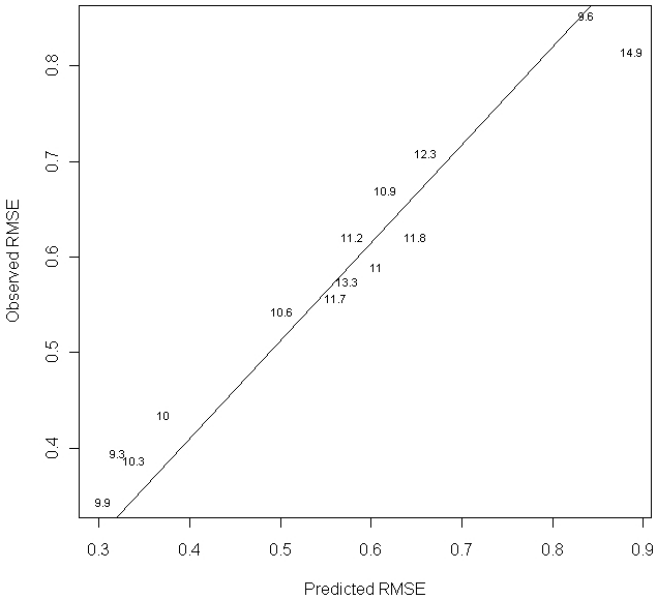


Fig. 2. Observed RMSE versus predicted RMSE for each of the 14 (X, y) pairs.

The main technical issue for MCMC is to assess convergence to the stationary distribution. In addition, any Bayesian approach must assess sensitivity to the prior probability distributions. For Example 1, both a proper (integrates to 1) diffuse prior for β_0 and β_1 (see Appendix A.4) and an improper prior give essentially the same results. For all priors considered, both the prior and of course then the posterior put zero probability on $\beta_1 < 0$.

The MCMC stragegy is simple to implement using for example metrop in the mcmc package in R. In addition, the strategy easily accommodates non-Gaussian likelihoods, and estimation errors in σ_x and in σ_y. As an example, the MCMC-based average RMSE over the 14 Ys is 0.60, which is close (within the repeatability due to a finite (10 000) number of simulations) to the estimated average RMSE in $\hat{Y}$ of 0.61 RMSE by EIP. The repeatability of 100 simulations using 100 000 observations from the estimated posterior distribution is approximately ±1% relative. Figure 2 plots the observed RMSE versus the MCMC-based prediction of RMSE for each of the 14 (x, y) pairs. Generally, MCMC can provide accurate uncertainty estimates for any quantity of interest for any sample size. The main caveats are that the user must check for convergence of the sampling chain to the correct

https://doi.org/10.1051/ijmqe/2012010 Published online by Cambridge University Press

posterior distribution, and the assumed probability model for the data must be approximately correct.

Note that the average RMSE across the 14 (x, y) pairs of MCMC-based approach (0.61) is slightly (and repeatably across sets of 10 000 simulations) less than that of the ML-based EIP approach (0.63). However, because of the strong reliance on the claimed values for σ_x and σ_y, the strong assumption of the linear relation in equation (1), and the assumption of well-behaved measurement errors, we do not claim that either MCMC or ML is preferred for the example considered.

Similarly to this MCMC-based numerical Bayesian approach, Wang and Iyer [7] presented a "fiducial inference" numerical option for estimating $\text{cov}(\hat{\beta}_0,\hat{\beta}_1)$ and for testing whether $\beta_0 = 1$ and $\beta_1 = 1$. In our example 1, there is also interest in whether $\beta_0 = 1$ and $\beta_1 = 1$ because x and y are two options for measuring the same temperature. The fiducial distribution is similar to the Bayesian posterior distribution, but it would be valuable to compare inference quality of these two numerical options. Here, the MCMC-based 95% credible interval for $\hat{\beta}_0$ is (–6.8, 1.6) (which includes 0) and for $\hat{\beta}_1$ is (0.8, 1.6) (which includes 1) so there is no evidence for a discrepancy between the X-based and Y-based temperature measurement.

6 Summary

This paper extends Reed [4] by providing approximate expressions for $\{\text{var}(\hat{\beta}_0)$, $\text{var}(\hat{\beta}_1)$, and $\text{cov}(\hat{\beta}_0,\hat{\beta}_1)\}$ of ML estimators. Equations (3), (8) and (9) provide the simple extension to WLS that can be used for the EIP approach. A numerical evaluation of Example 1 suggests that the approximate expressions are adequate for $\{\text{var}(\hat{\beta}_0)$, $\text{var}(\hat{\beta}_1)$, and $\text{cov}(\hat{\beta}_0,\hat{\beta}_1)\}$ of ML estimators, but that the estimated variance of $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\hat{X}$ can have less-than-adequate quality in small samples such as Example 1 with $n = 14$ (x, y) pairs.

When Example 1 was extended to 42 (x, y) pairs, all approximate expressions were fully adequate. Unfortunately, it is difficult to know what constitutes an adequately large sample because of dependence of approximation quality on multiple parameters including β_0, β_1, σ_{y_i}, and σ_{x_i}. Therefore, an MCMC-based alternative is also described and provided. In practice for any specific application, if EIP based on equations (6), (8) and (9) is applied, and if very high quality estimates of for $\{\text{var}(\hat{\beta}_0)$, $\text{var}(\hat{\beta}_1)$, and $\text{cov}(\hat{\beta}_0,\hat{\beta}_1)\}$ are needed, then simulations as described in Sections 3 and 4 are needed. Alternatively, if one opts for MCMC, then the user should apply convergence diagnostics to ensure the MCMC runs that purport to sample adequately from the posterior actually do. In short, applying EIP requires a bit more effort than does WLS whether one uses equations (8) and (9) or MCMC. Both options can be extended to more than one predictor X, but the associated inference quality needs to be studied.

Acknowledgements. The authors acknowledge the National Nuclear Security Administration, offices of NA and NE, including the Next Generation Safeguards Initiative, and the Los Alamos National Laboratory Directed Research Program.

Appendix

A.1 Expression for the ML estimates

For the case with σ_{x_i} and σ_{y_i} depending on the respective true values for observation i, Reed [5] extends Madansky [9], and Kulathinal et al. [17] provide an option involving the estimation-maximization algorithm. Kulathinal et al. [17] suggest the Fisher information as a basis to approximate $\text{cov}(\hat{\beta}_0,\hat{\beta}_1)$, which we have shown is adequate for Example 1, but not adequate as an intermediate step in estimating the observed average RMSE in $\hat{Y}$ (see Sect. 4).

We have verified that Reed's [5] more general expression gives the same estimators as ML. For reader convenience, Reed's [5] expression is given here. Define

$$\langle x\rangle = \frac{\sum_{i=1}^{n} W_i x_i}{\sum_{i=1}^{n} W_i} \text{ and } \langle y\rangle = \frac{\sum_{i=1}^{n} W_i y_i}{\sum_{i=1}^{n} W_i}.$$

Then define $U_i = x_i - \langle x\rangle$ and $\text{V}_i = y_i - \langle y\rangle$. Then $\hat{\beta}_1$ is the solution to the quadratic equation $A\hat{\beta}_1^2 + B\hat{\beta}_1 + C = 0$, where

$$A = \sum_{i=1}^{n} \frac{W_i^2 U_i V_i}{W(x_i)}, B = \sum_{i=1}^{n} W_i^2 \left\{ \frac{U_i^2}{W(y_i)} - \frac{V_i^2}{W(x_i)} \right\},$$

$$\text{and } C = \sum_{i=1}^{n} \frac{W_i^2 U_i V_i}{W(y_i)}.$$

A.2 Fisher information to approximate the covariance matrix of ML estimators

There is literature on using Fisher information to approximate the covariance matrix of ML estimators [18]. In equation (7), $S = \sum_{i=1}^{N} W_i \left(y_i - \beta_0 - \beta_1 x_i\right)^2$ is proportional to the likelihood, so derivatives with respect to β_0 and with respect to β_1 can be computed and the Fisher information is easily computed, and R code to do so is available upon request.

A.3 Prediction

Carroll et al. [1] state: "... if one wants to predict a response based on the error-prone version x of X then unless we want to develop a prediction model using data from one population and want to predict in another population,

https://doi.org/10.1051/ijmqe/2012010 Published online by Cambridge University Press

it rarely makes sense to worry about measurement error in x. The reason is simple: x is an error-free measurement of itself!"

To illustrate, we generated 100 X values on Uniform (0, 10) and corresponding y_{true} values using $y_{\text{true}} = \beta_0 + \beta_1 x_{\text{true}}$ with $\beta_0 = -2.312$ and $\beta_1 = 1.166$, which are the EIP-based estimates from Example 1 in Section 4. We set $\sigma_x = 4.4$ for all x and $\sigma_y = 8.6$ for all y. The resulting RMSE and average bias based on 10 000 simulations are 4.5 and ∼0, which are repeatable to within ±0.01 across sets of 10 000 simulations. We then changed the X values to lie on Uniform (0,15) and doubled the value of σ_x from 4.4 to 8.8 to represent "another population". The resulting estimated RMSE and average bias are then 6.1 and –1.3. If we know the second population will have $\sigma_x = 8.8$ and have X lie in (0,15), then a better estimation strategy can be developed.

A.4 R code for MCMC option

Here is R code for the MCMC option describe in Section 5.

```
ludv = function(vec,xmeas=xmeas,ymeas=ymeas,
sigmay=sigma.y, sigmax=sigma.x) {
# vec holds a,b,rep(mu1,n),sigmay where n is the number
of measured x,y pairs
n = length(xmeas)
# model: y ∼ N(mu,sigmay), xmeas ∼ N(xtrue,sigmax)
with mu = a + b (xtrue)
# logliklihood:
mu1 = vec[1:2] # a and b
mu2 =vec[3:(2+n)] # xtrue
theta = mu1[1] + mu1[2]*mu2
temp1 = sum(-1/2*(xmeas-mu2)^2/sigmax^2) +
sum(-1/2*(ymeas-theta)^2/sigmay^2)
# prior:
temp2 = 0
if(any(mu2 <0)) {temp2 = -100}
temp3 = temp1 + temp2
if(is.na(temp3)) {temp3 = -100}
if(temp3 == -Inf) {temp3 = -100}
return(temp3)
}
out= metrop(obj=ludv,initial=vec.init,nbatch=10000,
nspac=200,blen=1,scale=.075,
xmeas=xmeas,ymeas=ymeas,sigmay=sigmay,
sigmax=sigmax)
```

References

1. R. Carroll, D. Ruppert, L. Stefanski, C. Crainiceanu, *Measurement Error in Nonlinear Models* (Chapman & Hall, 2006)
2. W. Deming, *Statistical Adjustment of Data* (Wiley, 1943)
3. W. Fuller, *Measurement Error Models* (Wiley, 1987)
4. B. Reed, A spreadsheet for linear least-squares fitting with errors in both coordinates, Phys. Educ. **45**, 93−95 (2010)
5. B. Reed, Linear least-squares fits with errors in both coordinates. II: Comments on parameter variances, Am. J. Phys. **60**, 59−62 (1992)
6. D. York, N. Evensen, M. Martinez, J. Delgado, Unified equations for the slope, intercept, and standard errors of the best straight line, Am. J. Phys. **72**, 367−375 (2004)
7. C. Wang, H. Iyer, Fiducial approach for assessing agreement between two instruments, Metrologia **45**, 415−421 (2008)
8. M. Brown, Robust line estimation with errors in both variables, J. Am. Stat. Assoc. **77**, 71−79 (1982)
9. A. Madansky, The fitting of straight lines when both variables are subject to errors, J. Am. Stat. Assoc. **54**, 173−205 (1959)
10. W. Press, S. Teukolsky, W. Vettterling, B. Flannery, *Numerical Recipes, The Art of Scientific Computing*, 3rd edn. (Cambridge University Press, 2007)
11. J. Toivanen, J. Dobaczewski, M., Kortelainen, K. Mizuyama, Error analysis of nuclear mass fits, Phys. Rev. C **78**, 0343061−0343068 (2008)
12. R Development Core Team 2004 R: A Language and Environment for Statistical Computing
13. T. Burr, P. Knepper, A study of the effect of measurement error in predictor variables in nondestructive assay, Appl. Radiat. Isotopes **53**, 547−555 (2000)
14. C. Elster, B. Toman, Bayesian uncertainty analysis for a regression model versus application of GUM supplement 1 to the least-squares estimate, Metrologia **48**, 233−240 (2011)
15. W. Gilks, S. Richardson, D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman & Hall/CRC, 1996)
16. P. Dellaportas, D. Stephens, Bayesian analysis of errors-in-variables regression models, Biometrics **51**, 1085−1095 (1995)
17. S. Kulathinal, K. Kuulasmaa, D. Gasbarra, Estimation of an errors-in-variables regression model with the variance of the measurement errors vary between the observations, Stat. Med. **21**, 1089−1101 (2002)
18. M. DeGroot, *Probability and Statistics*, 2nd edn. (Addison-Wesley, 1986)

https://doi.org/10.1051/ijmqe/2012010 Published online by Cambridge University Press